CR-66-198-1

RESEARCH ON LINEAR SYSTEMS OF A VERY LARGE SIZE

SCIENTIFIC STATUS REPORT

for the period

6 May 1966 to 30 September 1966

24 October 1966

Prepared for

National Aeronautics and Space Administration
Washington, D. C.

Prepared by

Raytheon Company
Autometric Operation
4217 Wheeler Avenue
Alexandria, Virginia

1. INTRODUCTION

The investigation has proceeded along three directions. First, the topological characteristics of the matrix $\underline{A}$ in the equation

$$y = Ax$$

have been studied in a search for parameters which would be described in the discrete topology of A^T and which could be traced through the successive transformations:

$$B \equiv A^T A$$
$$N \equiv B^{-1}$$

A summary is given in Section 2 of the results to date.

Next, the properties of A, $A^T A$, and $(A^T A)^{-1}$ as continuous (infinite) matrices have been studied with an eye to relating the maxima and minima of A to those of $A^T A$ and $(A^T A)^{-1}$. This work is summarized in Section 3.

Finally, studies have been made of known procedures to see if these can be modified or new approaches found which will give solutions in a smaller number of steps. This work has led to derivation of two iterative procedures which are so different from each other that they are discussed separately in Sections 4 and 5. Section 6 is a bibliography supplementing that submitted with the proposal.

2. MATRIX TOPOLOGY

A study of the literature on the topology of matrices has not shown any results of particular use in the solution of large matrixes other than for reducing or separating matrices into decoupled submatrices. (Two submatrices are called "decoupled" if they do not contain index numbers in common.) A "decoupling tensor" can be defined which acts to transform an nxm matrix into a block diagonal square matrix nxn, but it does not seem to have much theoretical interest. Of more interest are a pair of parameters, which I call coupling factors, C^r_{ij} and C^c_{iy}, which are defined by:

$$C^r_{ij} = \sum_k \mu\ (a_{ik}, a_{jk}) \quad ,$$

where

$$\mu(z) = \begin{matrix} 1, & \nu = 0 \\ 0, & \nu \neq 0 \end{matrix}$$

and

$$C^c_{ij} = \sum_k \mu\ (a_{ki},\ a_{kj}) \quad .$$

These parameters are defined for rectangular as well as square matrices. They are invariant under permutation transformations and under a number of other operations. Certain functions of C^r_{ij} and C^c_{ij} are invariant under the transformation:

$$A^T A$$

and it is hoped that they can be traced further into $(A^TA)^{-1}$. They can be shown to be related to auto- and cross-correlation factors defined on continuous matrices.

$$\frac{\partial n_{kl}}{\partial b_{ij}} = - \quad n_{ki} \quad n_{jl} \qquad ,$$

where n_{kl} is an element of:

$$N \equiv B^{-1}$$

and b_{ij} is an element of:

$$B \equiv A^TA \qquad .$$

There is a path open for connecting the error ε_n in N through:

$$\varepsilon_n = \frac{\partial n_{kl}}{\partial b_{ij}} \quad \Delta b_{ij}$$

to the error Δb in b and to the coupling factors, either in the discrete case or in the continuous care or perhaps both.

Attention is being paid to the possibility connecting the coupling factors to Betti numbers and such like other topological parameters. They can be related to graphs and hence to incidence matrices, but the extent to which such relationships can be used is not known.

3. HANDLING LARGE MATRICES AS CONTINUOUS FUNCTIONS

This is an investigation of the possiblities of developing procedures analogous to matrix and vector product methods for piece-wise continuous functions of two variables. For this purpose the theory of convolution quotients will be used, since it gives a systematic way of dealing with the unit that must be used in an algebra where the products correspond to integrals. This theory is presented in "Operational Calculus and Generalized Functions: by Arthur Erdelyi, and is briefly given as follows.

A ring is defined whose elements are continuous functions of a non-negative real variable. Addition is defined as usual, but multiplication is defined as concolution:$^{-(1)}$i.e.,

$$f*g(t) = \int_0^t f\ (t-u)\ g(u)\ du \quad .$$

This multiplication is distributive, (with respect to Scalars) commutative, and associative. This ring C is then extended to a field F, much in the same way that the ring of integers is extended to the field of rationals, that is to say that pairs of elements of the ring are regarded as elements of the new field. (More precisely, equivalence sets of pairs are the elements in the same way that 6/9 and 4/6 are equivalent to each other, and the equivalence class containing them is represented by 2/3.) The pair (f,g) is equivalent to (a,b) if:

$$a*d = b*c$$

in the convolution sense. The equivalence class containing these is written a/b (or f/g). It can be shown that the field F (of convolution quotients) is a vector space and that it contains (in the sense of isomorphism) the real and complex numbers, the continuous functions in C, the (equivalence classes of) locally integratable functions, and the so-called impulse functions, one of which serves as the previous mentioned unit under convolution. This unit must correspond to the unit matrix.

The operations defined in the field F are:

$$\frac{a}{b} + \frac{c}{d} = \frac{a*d + b*c}{bd} \quad ; \quad \alpha \left(\frac{a}{b}\right) = \left(\frac{\alpha a}{b}\right) \quad ; \quad \left(\frac{a}{b}\right) \left(\frac{c}{d}\right) = \left(\frac{a*c}{b*d}\right) \quad .$$

In the cases we wish to pursue, the function is to be zero for all values of the arguments larger than a fixed number, probably 1; and consequently the convolutions can be regarded as, in the integral case, as regular intergrals with one of the factors, the mirror image of the function at issue. In the vector case this will entail dot products with vectors "put in backwards". Due to the nature of the convolution integral:

$$\int_0^t f\ (t-u)\ g(u)\ du = h\ (t)$$

the function (t-u) cannot be regarded as an arbitrarily chosen function of t and u, nor is the result h (t) strictly analogous to a dot product. With the interpretation of t as a constant (analogous to the size of the matrix problem in question) we get an analogue for the dot product. All the theory of convolution quotients holds for this interpretation, and in particular, the theory of convergence of convolution quotients, which is a generalization of the theory for scalars and of uniform convergence of continuous functions. That is, if a sequence of continuous functions converges uniformly on an interval in the usual sense, then they will converge in the convolution sense. (cf - Erdelyi, ref. 4)

$$\lim_{h\to\infty} \left\{ n\ f\ (nt) / \int_{0}^{\infty} [f(n)\ du] \right\} \to 1 \quad ,$$

where 1 is the unit under convolution or as it is sometimes called, the delta function. Notice the above functions do not converge uniformly in the usual sense.

In order to extend this analogy to the matrix case, it is necessary to examine the integrals of functions of two variables,

$$\int_{4=0}^{1} f(x,h)\ g(u,y)\ du = h(x,y) \quad .$$

For this we enlist the theory of integral equations of the first kind, and of Green's functions. (cf - Morse and Feshbach Sec. 8.3, ref. 3)

Transition from integrals to scalar products can be approximated with the formula:

$$\int_{a_n}^{a_{n+1}} f(x)\ dx = \lim_{i\to\infty} \sum_{i=0}^{\infty} \left(\frac{1}{(2_i+1)} \right) \left[f^{(2i)}\ \frac{a_n+a_{n+1}}{2} \right] \left(a_{n+1}-a_n \right)^{2i+1} \quad .$$

4. AN ITERATIVE SCHEME OF THE PROJECTION TYPE

This section deals with two matrix iterative methods which we believe to be original. The methods are projective, like tyose of Kacmarg and Cimmeno (see ref. 2), but are based on a non-standard geometric interpretation of a system of linear equations. The first method and its variants are iterative, while the second method is iterative direct, like that of conjugate gradients. Both ideas are in rough form and are in the process of being modified. There has been a limited amount of numerical testing.

The geometric picture arises by inverting the usual concept of solution spaces intersecting a point which solves the system. Instead, we have an "equation" space, a certain element of which can be easily modified to give the solution.

Let the system be:

$$BX = C$$

where

B has rows $B_i, \ldots, B_n$

$X = (x_j)$ is the solution, and

$C = (c_j)$ is the constant vector.

Divide each equation.

$$B_i X = C_i \tag{1}$$

by C_i. If C_i is zero, simply add another equation whose constant is non-zero to (1) before dividing.

The new system is where A has rows

$$A_i = B_i/C_i \qquad ,$$

and

$$1 = \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \qquad .$$

The rows of A can be thought of as n points in n space which define an n-h dimensional hyperplane H (if A has rank n, then k = 1). A point U of H is defined by:

$$\begin{aligned} U &= (\Sigma\, \alpha_i\, A_i)\ x \\ &= \Sigma\, \alpha_i \qquad . \\ &= 1 \end{aligned}$$

Thus, any n points in H correspond to an nxn matrix M, say, such that MX = 1. If the points are vertices of a convex set, them M is non-singular.

Let P be the point in H such that the vector P is normal to H. Every row A_i of A can be expressed as:

$$A_i = P + S_i \qquad ,$$

where S_i is a vector in H. We have A_i P = P for all i.

In matrix form,

$$AP = \begin{bmatrix} P^2 \\ P^2 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \quad .$$

thus,

$$X = P/P^2 \quad .$$

The first method proceeds as follows. We choose some initial approximation X_o (Example: the bary center of the A_i) and some initial point A_k. The line

$$U = A_k + \theta\ (X_o - A_k) \tag{3}$$

obviously has in H. We wish to find the point U closest to P. Since this point is also closest to the origin we have:

$$U\ (X_o - A_k) = \theta$$

$$\theta_1 = A_k\ (X_o - A_k)/(X_o - A_k)^2 \tag{4}$$

$$X_1 = A_k + \theta_1\ (X_o - A_k) \tag{5}$$

θ cannot be 1 for all A_i. If it were, then all vectors $X_o - A_i$ would be normal to X_o-P. Thus, all points A_i would be continued in an n-h-1 dimensional subspace, which is impossible because H has dimension h-k.

To continue, we repeat the process with a new A_i to get X_2 and so on until we have used all of the A_i. Then we start over and continue until the desired accuracy is achieved. Since quantities from i through 1 iteration can be used, the number of multiplications in the ith iteration is 2n.

The process converges for each test matrix used including 2 singular ones.

For a variation of the method X_i^2 can be minimized as a function of X_{i-1}. This process also requires 2n multiplications per iteration, i.e. to compute X_{i-1}. X_i is then computed so as to minimize X_{i+1}^2. In each case tested there was improvement, sometimes considerable, over the first process. However, initializing requires n^2 more multiplications and another column of storage. In general, X_i^2 can be minimized as a function of the proceding k approximations. An additional kn^2 multiplication to initialize the k columns of storage are required over the first process. These further refinements were not tested.

For the iterative direct process we proceed as follows. Assume that A is non-singular. Choose H_k which is closest to P, i.e., has minimum length. Choose A_e such that, using A_a as X_o, the point X_1 is closer to P than for other A_i. We will iteratively "correct" the vector $G = A_h - A_\ell$ until it has the same direction as $P-A_k$.

Any point U in the n-2 dimensional hyperplane containing A_k whose normal is $P-A_k$ satisfies

$$A_k U = A_k^2 \quad . \tag{6}$$

If we set

$$U_i = A_i + \theta_i \ (A_i - A_\ell) \qquad ,$$

and choose θ_i so that U_i satisfies (6), then we can project the A_i exclusive of A_k and A_ℓ onto the hyperplane. Thus, we have n-1 points in a n-2 dimensional hyperplane. We can form a basis, say V_1, $V_2, \ldots, V_{n-2}$, which can be orthogonalized one vector at a time. If the new basis is $\alpha_1, \ldots, \alpha_{n-2}$ then we can make G orthogonal to it by well known formulas. We can create the α_i one at a time and correct G each. Once G is corrected, then obviously P lies on the line $U = A_k + \theta G$. The process requires on the order of n^3 multiplications. However, the first iteration requires only 6n multiplications, the second 7n, the third 9n, and so on adding 2n each time. The last iteration requires about $2n^2$ multiplications. Thus, most of the work is concentrated in the last few iterations. This may be an advantage over the method of conjugate gradients, which requires n^2 multiplications for each iteration. Numerical experience shows that the method does provide an exact solution.

5. AN ITERATIVE SCHEME OF THE SEPARATION TYPE

The standard relaxation method can be written as:

$$x^{(n+1)} = \omega_n D^{-1} x^{(n+1)} - (\omega_n - 1 - \omega_n D^{-1} U) \Delta\alpha^{(n)} + \omega_n D^{-1} y \qquad (1)$$

where D, U, and L are the strictly diagonal, upperdiagonal, and lower diagonal components of the nxn (not necessarily symmetric) matrix B. ω_n is the relaxation factor appropriate to the nth step, and distinguishes the various common relaxation methods. E.g, for $\omega_n \equiv 1$, equation 5.1 characterizes the Gauss Seedel method while for ω_n = a constant >1, or <1 we have the over-relaxation or under relaxation method, respectively.

Now the over relazation method can be shown (ref.) to give very rapid convergence, on the average and after a very large number of steps, for certain types of B matrix, and in fact for some types of B matrix it converges faster than any other iterative procedure. It does not converge faster for many types of matrices and in fact, even for B matrices of the type mentioned it may often converge more slowly than other methods for the first few steps. A procedure called the Chebychev semi-terative method (ref. 1) gives rapid local convergence for may types of B matrix but suffers from a number of disadvantages, one of which is the requirement that the spectral radius of B be known, at least approximately. To avoid these difficulties and to get still more rapid convergence, a method which requires no knowledge of B's eigen values and which makes fullest use of available information is being studied. The prototype

REFERENCES

1. 1962 Varga, R., Matrix Iterative Analysis, Prentice-Hall, Inc., Englewood Cliffs, N.J., pp. 97-154.

2. 1959 Bodewig, E., Matrix Calculus, North-Holland Publishing Company, Amsterdam Interscience Publishers, Inc., N.Y., pp. 186-188.

3. 1953 Morse, P. M. and Feshbach, H., Methods of Theoretical Physics, Vol. I, Sect. 8.3

4. 1962 Erdelyi, Operational Calculus and Generalized Functions, Holt, Reinhardt and Winston, New York